The stochastic quasi-steady-state assumption: Reducing the model but not the noise

Rishi Srivastava,[1,a)] Eric L. Haseltine,[2,b)] Ethan Mastny,[1,c)] and James B. Rawlings[1,d)]
[1]*Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Wisconsin, 53706-1607, USA*
[2]*Division of Chemistry & Chemical Engineering 210-41, California Institute of Technology, Pasadena, California 91125, USA*

(Received 26 October 2010; accepted 24 March 2011; published online 20 April 2011)

Highly reactive species at small copy numbers play an important role in many biological reaction networks. We have described previously how these species can be removed from reaction networks using stochastic quasi-steady-state singular perturbation analysis (sQSPA). In this paper we apply sQSPA to three published biological models: the pap operon regulation, a biochemical oscillator, and an intracellular viral infection. These examples demonstrate three different potential benefits of sQSPA. First, rare state probabilities can be accurately estimated from simulation. Second, the method typically results in fewer and better scaled parameters that can be more readily estimated from experiments. Finally, the simulation time can be significantly reduced without sacrificing the accuracy of the solution. © *2011 American Institute of Physics*. [doi:10.1063/1.3580292]

I. INTRODUCTION

Intrinsic noise caused by discrete numbers of reacting molecules is an inherent feature of many cellular processes.[1–5] How individual cells control and exploit intrinsic noise has been an area of active research for the past decade. One useful tool for understanding this noise is the chemical master equation, which models a reacting system as a discrete Markov process in which jumps from one discrete state to another represent chemical reactions. The solution of the master equation is computationally tractable only for simple systems. Rather, approximation techniques such as finite state projections[6] or stochastic simulation algorithm (SSA)[7,8] are employed to reconstruct the probability distribution and its statistics (usually the mean and variance). Applying these techniques to solve models of biological processes leads to significant improvements in understanding of intrinsic noise and its effect on cellular behavior. For example, Arkin and McAdams showed that in the bacteriophage λ infection, intrinsic noise could bifurcate identically infected cells to either dormant (lysogenic) or reproductive (lytic) states.[9] Vilar *et al.* demonstrated that biochemical oscillators could still reliably function even in the presence of such noise.[10] Weinberger *et al.* showed (both computationally and experimentally) that noise could generate transient bursts of activity in HIV-1 Tat transactivation, which contributes to latency in HIV-1.[11] Hensel *et al.* performed computational studies to illustrate the dynamics of an intracellular infection process demonstrating coupling between a highly reactive species and a rapidly increasing species.[12]

While these works are clearly successes of the modeling community in understanding the intricacies of intrinsic noise, the reality is that stochastic models of biologically relevant systems remain difficult to construct and even more difficult to understand. Many of these difficulties stem primarily from the significant cost of solving these models, particularly when using SSA simulation; see Gillespie[13] for a recent review of progress made on this front. Moreover, for many systems biology models, the available experimental measurements are not sufficient to confidently estimate the parameters of the model.[14] While these models can still make well-constrained predictions, the unnecessarily large number of parameters in these models increases the cost of calculating parameter sensitivities, i.e., how the model predictions change with perturbations to the parameters. Often these large number of parameters lead to model stiffness, a phenomenon in which significant computational expense is incurred in simulating some subset of the reactions in the model.

Stiffness can arise when some set of reversible reactions occur much more frequently than the remaining reactions, and the affected species typically remain at reasonable (nonzero) numbers. Such stiffness is analogous to the deterministic concept of reaction equilibrium and has been the subject of several recent studies.[15–21] A separate but equally important source of stiffness results from highly reactive species. This phenomenon results when reaction intermediates, known as quasi-steady-state (QSS) species, react so rapidly that their average number throughout the simulation is nearly zero or their average number is much smaller than the other species (the reactants and products).[22,23] Rao and Arkin hypothesized that the stochastic quasi-steady-state reduction should lead to the same reduced model as the deterministic quasi-steady-state reduction. Mastny *et al.*,[23] however, presented counterexamples in which the reduced models from deterministic and stochastic quasi-steady-state reduction are different. Using singular perturbation analysis, Mastny *et al.*[23] demonstrated that the QSS species can be removed from the master equation to yield a reduced

[a)]Electronic mail: rsrivastava2@wisc.edu.
[b)]Electronic mail: haseltin@caltech.edu.
[c)]Electronic mail: ethanmastny@gmail.com.
[d)]Electronic mail: rawlings@engr.wisc.edu.

 © 2011 American Institute of Physics

Downloaded 17 May 2011 to 131.215.220.185. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

master equation for the remaining species. They termed this method stochastic QSS singular perturbation analysis (sQSPA). Inspection of the reduced master equation yields reduced reaction expressions, reduced stoichiometries and fewer parameters in comparison to the original master equation. Notably, the reduced reaction expressions do not always correspond to those obtained by using the traditional deterministic quasi-steady-state analysis. Another source of stiffness can occur when there are two distinct type of species present in the system: a highly reactive species and a rapidly increasing species. This stiffness causes small time steps in SSA and leads to significant slow down of SSA simulations. For this stiffness, we use a variant of sQSPA, stochastic QSS singular perturbation analysis with Ω expansion (sQSPA-Ω).

In this paper, we demonstrate the utility of sQSPA and sQSPA-Ω for simulating and understanding stochastic reaction models. We choose two previously published models from the literature that appear to have highly-reactive, or QSSA species: *the pap operon regulation*[6] and *a biochemical oscillator.*[10] We also consider a simple system, *a fast fluctuation*, which contains a highly reactive species coupled to a rapidly increasing species. This coupling of highly reactive species with rapidly increasing species causes a large computational load for SSA simulations.[12] By reducing these three models using sQSPA and sQSPA-Ω, we show that we can lower the model complexity without altering the inherent noise characteristics of the full model. The reduced model for the *pap operon regulation* is also useful in estimating a rare state probability from simulation. This rare state probability is not accurately determined by either direct SSA simulation or Kuwahara and Mura's[24] recent general purpose method for estimating rare state probabilities in stochastic kinetic models. The reduced model for the *biochemical oscillator* leads to simplification in the parameter estimation problem as well as significant reduction in the simulation time. The reduced model for the *fast fluctuation* problem also significantly reduces the simulation time.

II. RESULTS

Example 1: Pap operon regulation. Here we consider the gene state switch model of the Pyelonephritis-associated pili (Pap) regulatory network considered by Munsky *et al.*[6] This model describes the states (g_1 to g_4) of the pap operon as a function of time. The schematic of four possible states and the mode of transition among them is shown in Fig. 1. The reaction stoichiometries and time invariant rates of transition (r_1 to r_8) are given in Table I.

The master equation for the system is

$$\frac{dP_1}{dt} = -(r_1 + r_3)P_1 + r_2P_2 + r_4P_3, \quad (1)$$

$$\frac{dP_2}{dt} = -(r_2 + r_5)P_2 + r_1P_1 + r_6P_4, \quad (2)$$

$$\frac{dP_3}{dt} = -(r_4 + r_7)P_3 + r_3P_1 + r_8P_4, \quad (3)$$

$$\frac{dP_4}{dt} = -(r_6 + r_8)P_4 + r_5P_2 + r_7P_3, \quad (4)$$

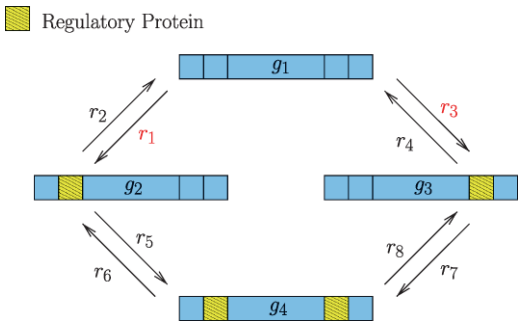

FIG. 1. Schematic diagram of the Pap regulatory network. There are four possible states of the pap operon depending on the LRP-DNA binding.

in which P_i: $i = 1, 2, 3, 4$ is the probability of state g_i and r_j: $j = 1, 2, \ldots, 8$ is the rates of transition defined in Table I. For the rates specified in Table I, r_1 and r_3 are large compared to all other rates. We apply the previously described sQSPA reduction technique.[23] In the supporting information,[25] we show that Eqs. (1)–(4) can be reduced appropriately for the two time regimes: the fast time scale and the slow time scale. The fast time scale regime occurs initially for a short period of time. During the fast time scale regime, probabilities of all the states directly affected by r_1 and r_3 change rapidly. The slow time scale follows this fast time scale regime. From Fig. 1, we can see that the only state that is not directly affected by r_1 and r_3 is state g_4.

On the fast time scale we denote the probability of state g_i as $\hat{P}_i$ and use the following power series expansion for $\hat{P}_i$

$$\hat{P}_i = \hat{W}_{i0} + \epsilon\hat{W}_{i1} + \epsilon^2\hat{W}_{i2} + O(\epsilon^3),$$

in which $\hat{W}_{ij}$ is the jth-order approximation of probability of g_i and ϵ is a small parameter. In the supporting information we show that for the fast time scale, the probability of different states is given by

$$\hat{W}_{10} = e^{-\tau}, \quad (5)$$

$$\hat{W}_{20} = K_1(1 - e^{-\tau}), \quad (6)$$

$$\hat{W}_{30} = K_3(1 - e^{-\tau}), \quad (7)$$

$$\hat{W}_{40} = 0, \quad (8)$$

in which $\tau = t(r_1 + r_3)$ is the rescaled time. Equation (5) indicates that as τ tends to ∞, i.e., we approach the boundary

TABLE I. Reaction stoichiometry and reaction rates for pap operon regulation.

Number	Reaction stoichiometry	Reaction rate (r_i)
1	$g_1 \longrightarrow g_2$	100.
2	$g_2 \longrightarrow g_1$	0.625
3	$g_1 \longrightarrow g_3$	100.
4	$g_3 \longrightarrow g_1$	1.033
5	$g_2 \longrightarrow g_4$	0.99
6	$g_4 \longrightarrow g_2$	1.033
7	$g_3 \longrightarrow g_4$	0.99
8	$g_4 \longrightarrow g_3$	0.625

Downloaded 17 May 2011 to 131.215.220.185. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

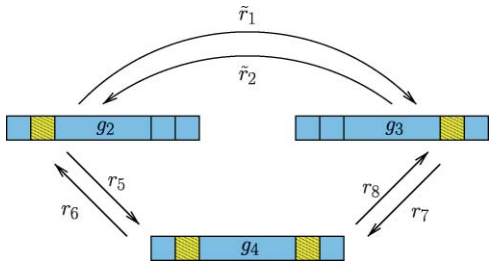

FIG. 2. Reduced system in the slow time scale regime.

between the fast time scale and the slow time scale regimes, the probability of state g_1 becomes small.

For the slow time scale we use following power series expansion for P_i

$$P_i = W_{i0} + \epsilon W_{i1} + \epsilon^2 W_{i2} + O(\epsilon^3),$$

in which W_{ij} are the jth-order probabilities of state g_i: $i = 1, 2, 3, 4$. In the supporting information we show that in the slow time scale regime

$$W_{10} = 0, \tag{9}$$

$$\frac{dW_{20}}{dt} = -[\tilde{r_1} + r_5]W_{20} + \tilde{r_2}W_{30} + r_6 W_{40}, \tag{10}$$

$$\frac{dW_{30}}{dt} = -[\tilde{r_2} + r_7]W_{30} + \tilde{r_1}W_{20} + r_8 W_{40}, \tag{11}$$

$$\frac{dW_{40}}{dt} = -[r_6 + r_8]W_{40} + r_5 W_{20} + r_7 W_{30}, \tag{12}$$

in which $\tilde{r_1} = r_2 r_3/(r_1 + r_3)$ and $\tilde{r_2} = r_1 r_4/(r_1 + r_3)$ are both $O(1)$ constants. The reduced reaction mechanism in the slow time scale regime is shown in Fig. 2. Note that state g_1 is removed in the slow time scale regime. In Fig. 3, we show a comparison of probabilities of different states at time $t = 10s$ as obtained from the master equation solution of the full model, Eqs. (1)–(4) and from the $O(1)$ approximation of the

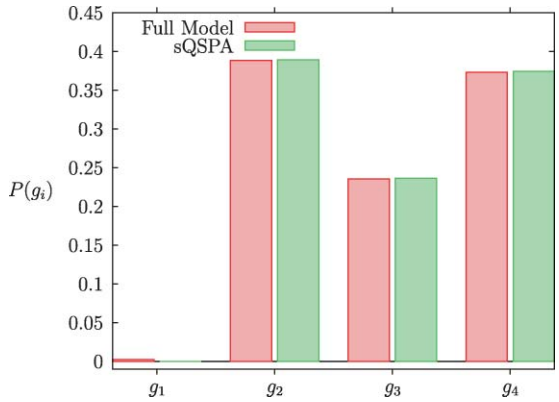

FIG. 3. Comparison of full model and sQSPA slow time scale reduced model at $t = 10$ s.

slow time scale solution of sQSPA, Eqs. (9)–(12). We can see that the $O(1)$ approximation of the slow time scale solution of sQSPA gives the probability of state g_1 as 0. This approximation is validated by the small value of the probability of state g_1 given by the master equation solution of the full model.

Increasingly better approximations of the probability of state g_1 can be obtained by higher-order terms in the sQSPA reduction.[23] In the supporting information we show that in the slow time scale regime, a better approximation of state g_1 is given by

$$\tilde{P}_1 = W_{10} + \epsilon W_{11} = \epsilon(r_2 W_{20} + r_4 W_{30}), \tag{13}$$

where $\epsilon = 1/(r_1 + r_3)$. Next we compare the probability estimate of state g_1 from the full model, Eqs. (1)–(4) and sQSPA slow time scale reduced model, Eq. (13). The result is shown in the left side of Fig. 4. The comparison demonstrates that after a small initial time interval, the master equation solution of both the full model and the sQSPA slow time scale reduced model of state g_1 are in good agreement. Next we show that sQSPA is useful for rare state probability estimation. In the right side of Fig. 4, we compare the probability estimate

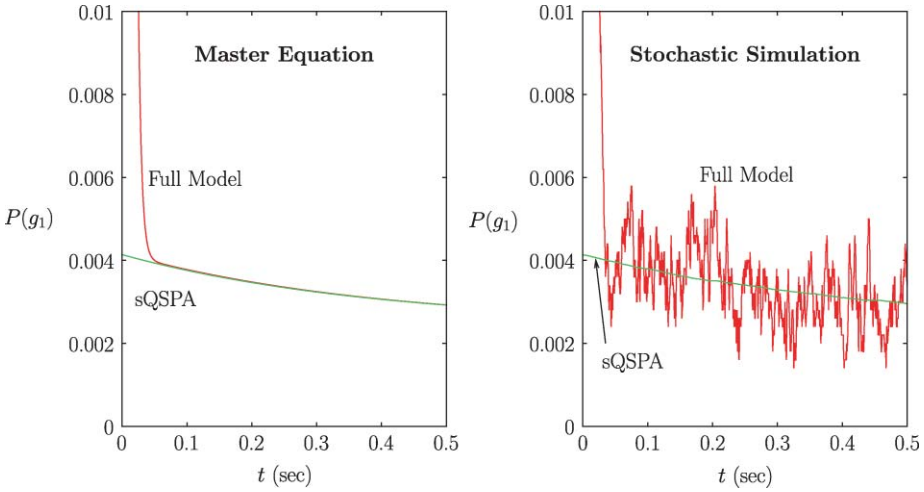

FIG. 4. A comparison of the full model and sQSPA reduced model. The left side shows the solution of the master equation for the full model and sQSPA-slow time scale reduced (sQSPA-stsr) model. The right side shows the estimate of the probability of state g_1 from full model SSA simulation and from sQSPA-stsr model SSA simulation.

Downloaded 17 May 2011 to 131.215.220.185. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

of state g_1 from 5000 SSA simulations of the full model, Eqs. (1)–(4) and from 5000 SSA simulations of sQSPA slow time scale reduced model, Eqs. (10)–(12). From simulations of the sQSPA slow time scale reduced model, we obtain values of W_{20}, W_{30}, and W_{40}. An estimate of rare state g_1 is then obtained from Eq. (13). The right side of Fig. 4 shows that the estimate of rare state g_1 is noisy and inaccurate from full model SSA simulations with 5000 samples, while this estimate is smooth and better with sQSPA slow time scale reduced model SSA simulations. Accurate estimation of rare state probabilities is a recurrent problem in stochastic simulation because this estimation requires large numbers of SSA simulations. The analysis of sQSPA enables these rare state probabilities to be determined from easily estimated probabilities of the nonrare states [e.g., use of Eq. (13)].

It is also interesting to compare this analysis to general purpose methods for estimating rare state probabilities in stochastic kinetic models. Kuwahara and Mura[24] present a general method to calculate rare state probabilities using the idea of importance sampling. Their method was subsequently improved by Roh *et al.*[26] employing a state dependent importance sampling parameter. Kuwahara and Mura's method relies on increasing the sampling of the rare state. The rate constants causing the rare states to have low probability are first biased (changed) so that the rare events occur more frequently in simulation; then the importance sampling weights are adjusted to remove the bias introduced by changing the rate constants. We implemented Kuwahara and Mura's method on the pap operon example by decreasing by a factor of 1000 both rate constants r_1 and r_3 for leaving state g_1. Surprisingly, the rare state probability estimate with this change in rate constants turns out to be *no better* than the full SSA implementation. The reason for this counter-intuitive lack of improvement has to do with the nature of the rare event. In the pap operon example, the rare state g_1 is visited *frequently*, but the *duration* in this state is small. But the decrease in r_1 and r_2 has no effect on the simulation because the time for the next event is chosen using the *original* rate constants, not the biased ones. So the duration in rare state g_1 is unaffected. The modified simulation does not visit state g_1 more frequently, and it does not remain in this state longer than in the original SSA simulation. Similarly, increasing r_2 and r_4 also does not give better estimates.

Example 2: Biochemical Oscillator. Sustained oscillations play a key role in biological processes such as circadian rhythms and cell-cycle dynamics. In many cases, these oscillations are driven by underlying biochemical reactions. Noise is an inherent component of cellular processes that causes cell-to-cell variation,[1–5] and is a potentially destabilizing influence that must be overcome to achieve sustained, reproducible oscillations. Vilar *et al.* examined this phenomenon using a model system,[10] and we reconsider the same reaction system

$$\begin{aligned}
\mathrm{A} + \mathrm{D_A} &\underset{\theta_A}{\overset{\gamma_A}{\rightleftharpoons}} \mathrm{D'_A},\\
\mathrm{D_A} &\xrightarrow{\alpha_A} \mathrm{D_A} + \mathrm{M_A},\\
\mathrm{D'_A} &\xrightarrow{\alpha'_A} \mathrm{D'_A} + \mathrm{M_A},\\
\mathrm{M_A} &\xrightarrow{\delta_{MA}} \text{degraded},\\
\mathrm{M_A} &\xrightarrow{\beta_A} \mathrm{M_A} + \mathrm{A},\\
\mathrm{A} &\xrightarrow{\delta_A} \text{degraded},\\
\mathrm{A} + \mathrm{D_R} &\underset{\theta_R}{\overset{\gamma_R}{\rightleftharpoons}} \mathrm{D'_R},\\
\mathrm{D_R} &\xrightarrow{\alpha_R} \mathrm{D_R} + \mathrm{M_R},\\
\mathrm{D'_R} &\xrightarrow{\alpha'_R} \mathrm{D'_R} + \mathrm{M_R},\\
\mathrm{M_R} &\xrightarrow{\delta_{MR}} \text{degraded},\\
\mathrm{M_R} &\xrightarrow{\beta_R} \mathrm{M_R} + \mathrm{R},\\
\mathrm{R} &\xrightarrow{\delta_R} \text{degraded},\\
\mathrm{A} + \mathrm{R} &\xrightarrow{\gamma_C} \mathrm{C},\\
\mathrm{C} &\xrightarrow{\delta_A} \mathrm{R}.
\end{aligned} \tag{14}$$

For the parameters given in Table II, the mRNA species M_A is a QSS species. A full SSA simulation of the model demonstrates that this species predominantly samples [0, 1] over the course of a typical simulation. To understand how one can reduce the model to eliminate this species, consider the simpler reaction system

$$\begin{aligned}
\mathrm{D_A} &\xrightarrow{\alpha_A} \mathrm{D_A} + \mathrm{M_A},\\
\mathrm{D'_A} &\xrightarrow{\alpha'_A} \mathrm{D'_A} + \mathrm{M_A},\\
\mathrm{M_A} &\xrightarrow{\delta_{MA}} \text{degraded},\\
\mathrm{M_A} &\xrightarrow{\beta_A} \mathrm{M_A} + \mathrm{A}.
\end{aligned} \tag{15}$$

TABLE II. Parameters for the biochemical oscillator. Initial conditions for unreported species are zero.

Parameter	Value	Units
γ_A	1.	$(\text{mol h})^{-1}$
θ_A	50.	h^{-1}
α_A	50.	h^{-1}
α'_A	500.	h^{-1}
δ_{MA}	1000.	h^{-1}
β_A	5000.	h^{-1}
δ_A	1.	h^{-1}
γ_R	1.	$(\text{mol h})^{-1}$
θ_R	100.	h^{-1}
α_R	0.01	h^{-1}
α'_R	50.	h^{-1}
δ_{MR}	50.	hr^{-1}
β_R	500.	h^{-1}
δ_R	0.2	h^{-1}
γ_C	2.	$(\text{mol h})^{-1}$
$D_A(t=0)$	1.	mol
$D_R(t=0)$	1.	mol

Downloaded 17 May 2011 to 131.215.220.185. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

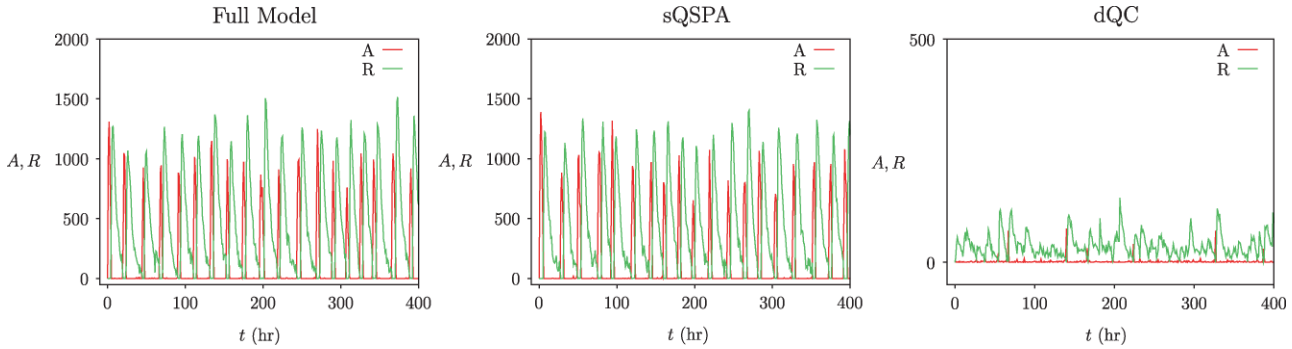

FIG. 5. Stochastic simulation of the biochemical oscillator. Proteins A and R vs time for the full model (left), sQSPA-reduced model (center), and dQC reduced model (right).

In the supporting information, we show that when $\beta_A, \delta_{MA} \gg \alpha_A, \alpha'_A, \delta_A$ this simpler system (15) is well approximated by

$$\begin{aligned}
&\mathrm{D_A} \longrightarrow \mathrm{D_A} + \mathrm{A} \qquad r_A^1 = \bar{\beta}_A(\bar{\delta}_{MA}\alpha_A D_A),\\
&\mathrm{D_A} \longrightarrow \mathrm{D_A} + 2\mathrm{A} \qquad r_A^2 = \bar{\beta}_A^2(\bar{\delta}_{MA}\alpha_A D_A), \qquad (16)\\
&\qquad\qquad \vdots\\
&\mathrm{D_A} \longrightarrow \mathrm{D_A} + n\mathrm{A} \qquad r_A^n = \bar{\beta}_A^n(\bar{\delta}_{MA}\alpha_A D_A),\\
&\mathrm{D'_A} \longrightarrow \mathrm{D'_A} + \mathrm{A} \qquad r_A'^1 = \bar{\beta}_A(\bar{\delta}_{MA}\alpha'_A D'_A),\\
&\mathrm{D'_A} \longrightarrow \mathrm{D'_A} + 2\mathrm{A} \qquad r_A'^2 = \bar{\beta}_A^2(\bar{\delta}_{MA}\alpha'_A D'_A), \qquad (17)\\
&\qquad\qquad \vdots\\
&\mathrm{D'_A} \longrightarrow \mathrm{D'_A} + n\mathrm{A} \qquad r_A'^n = \bar{\beta}_A^n(\bar{\delta}_{MA}\alpha'_A D'_A),
\end{aligned}$$

in which the M_A species no longer appears and $\bar{\beta}_A = \beta_A/(\delta_{MA} + \beta_A)$, $\bar{\delta}_{MA} = 1 - \bar{\beta}_A$. In the reduced system, the DNA species D_A and D'_A act as catalysts for production of the A protein. This system behaves similar to the catalytic example previously reported in Ref. 23 and is simulated in an analogous fashion. Namely, the sQSPA reduction indicates that the sum of all catalytic reactions involving D_A is

$$r_A^{\Sigma} = \sum_{j=1}^{\infty} r_A^j = \bar{\beta}_A \alpha_A D_A,$$

and the sum of all catalytic reactions involving D'_A is

$$r_A'^{\Sigma} = \sum_{j=1}^{\infty} r_A'^j = \bar{\beta}_A \alpha'_A D'_A.$$

Therefore, the catalytic production of A from D_A, catalytic reaction set (16) and D'_A, catalytic reaction set (17) can each be treated as a single reaction with rate r_A^{Σ} or $r_A'^{\Sigma}$. If the catalytic reaction set involving D_A is selected in the reduced model simulation, then one can use individual catalytic rates (the r_A^j's) to determine which catalytic reaction gets fired from the catalytic reaction set (16). A similar approach can be used if the catalytic reaction set involving D'_A gets selected.

Next we show that applying the stochastic and deterministic quasi-steady-state model reductions can lead to different reduced models. In the supporting information we show that if we apply deterministic quasi-steady-state classical dQC to the same reaction network (15), we get the following reduced mechanism:

$$\mathrm{D_A} \xrightarrow{\alpha_A} \mathrm{D_A} + A \qquad r_A = \bar{\beta}_A \alpha_A D_A, \qquad (18)$$

$$\mathrm{D'_A} \xrightarrow{\alpha'_A} \mathrm{D_A} + A \qquad r'_A = \bar{\beta}_A \alpha'_A D'_A. \qquad (19)$$

The sQSPA reduced model of the biochemical oscillator replaces the reaction set (15) in the full model (14) with reactions sets (16) and (17). The dQC reduced model of the biochemical oscillator replaces the reaction set (15) in the full model (14) with reactions sets (18) and (19). The reduced biochemical oscillator model as obtained by sQSPA and as obtained by dQC reveals important features of the two reductions. Simulating the reduced system obtained from sQSPA reproduces the oscillatory characteristic of the full model (Fig. 5). However, the reduced system obtained from dQC is grossly inaccurate and completely fails to reproduce oscillatory characteristics of the full model (Fig. 5). The reason that dQC reduction fails to reproduce the oscillatory characteristics of the full model is that dQC reduction underestimates the rate of production of A compared to both the full model and the sQSPA reduced model.

To more quantitatively compare the three different models—full, sQSPA reduced and dQC reduced, we compare expected value ($E[A(t)]$) and standard deviation ($\sigma[A(t)]$) of population of species A as obtained by the three models. To obtain the estimates of expected values and standard deviations at different time instants, we perform 3000 SSA simulations of each of the three models. To obtain $E[A(t)]$

TABLE III. Comparison of full, sQSPA reduced and dQC reduced models: $E[A(t)]$ and $\sigma[A(t)]$ obtained by sQSPA reduced model are close to those obtained by Full model. However, $E[A(t)]$ and $\sigma[A(t)]$ obtained by dQC reduced model is significantly different from those obtained by full model.

	$E[A(t)]$			$\sigma[A(t)]$		
t in h	Full	sQSPA	dQC	Full	sQSPA	dQC
0	0	0	0	0	0	0
5	192.77	195.08	1.13	208.21	203.34	4.58
10	0.17	0.14	1.30	0.72	0.90	4.90
35	14.36	14.11	1.66	96.86	95.90	5.13
100	93.65	96.31	1.69	249.47	253.46	5.32
200	129.80	117.43	1.73	287.08	271.38	5.88
400	125.69	127.49	1.82	282.08	287.35	5.88

Downloaded 17 May 2011 to 131.215.220.185. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

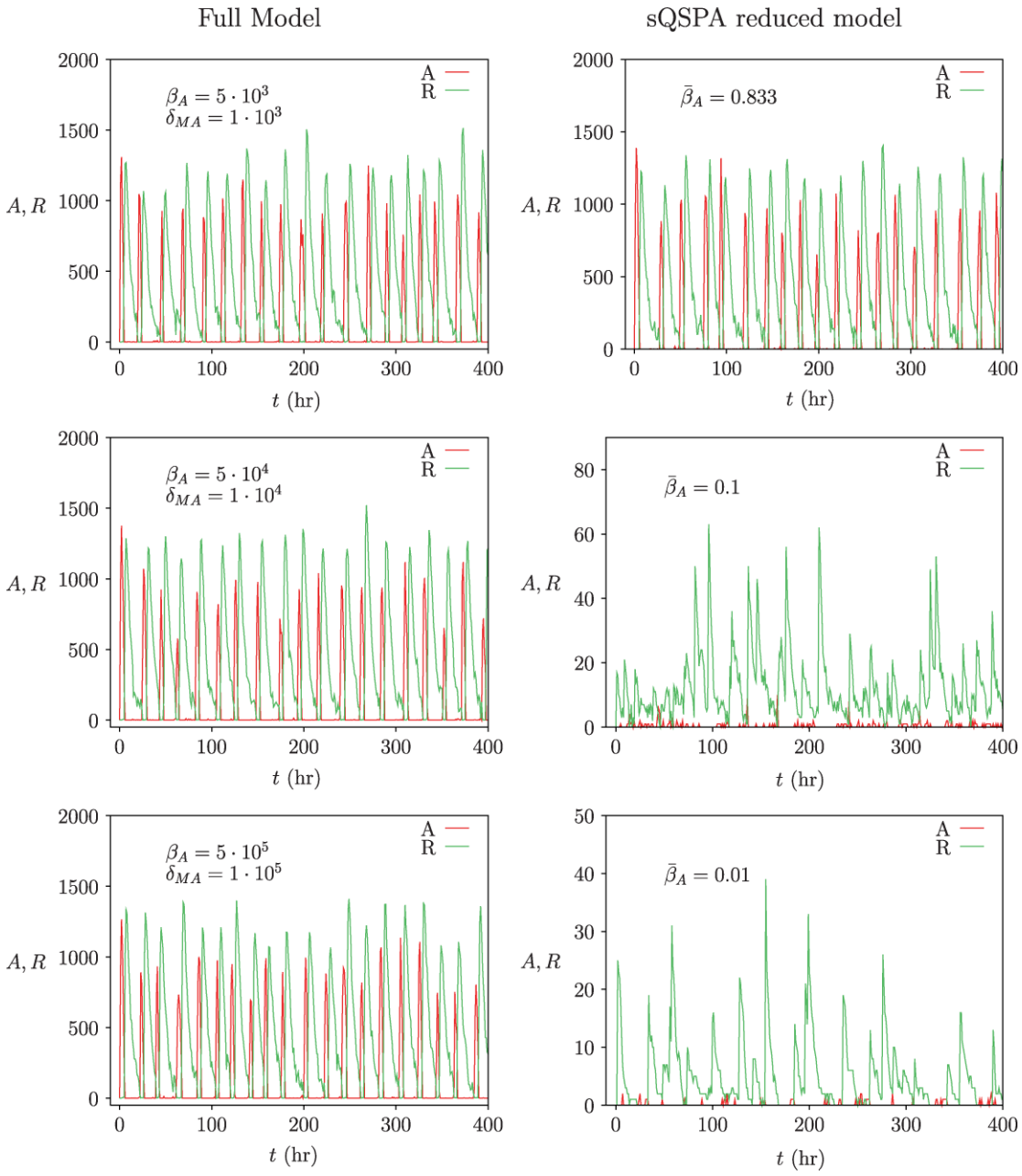

FIG. 6. Sensitivity of the full and sQSPA reduced models to parameters. Left column: Output of the full model for different value of β_A and δ_{MA} while keeping their ratio fixed; Right column: Output of sQSPA reduced model for various value of $\bar{\beta}_A$.

or $\sigma[A(t)]$ for a particular model, we use the following formulas:

$$E[A(t)] = \frac{1}{N_{\text{sim}}} \sum_{j=1}^{N_{\text{sim}}} A(t)$$

$$\sigma[A(t)] = \frac{1}{N_{\text{sim}} - 1} \sum_{j=1}^{N_{\text{sim}}} (A(t) - E[A(t)])^2$$

and $N_{\text{sim}} = 3000$ is the total number of SSA simulations performed. In Table III we see that at any given time $E[A(t)]$ and $\sigma[A(t)]$ as obtained by sQSPA reduced model are close to those obtained by the full model. However, the estimates of $E[A(t)]$ and $\sigma[A(t)]$ obtained with the dQC reduced model deviate significantly from those obtained by the full model.

The sQSPA model reduction changes the full model parameter estimation problem of two large parameters (β_A, δ_{MA}) into the estimation of one well-scaled parameter $(\bar{\beta}_A)$.[27] The left column of Fig. 6 demonstrates how the full model simulations change for different values of β_A and δ_{MA}. Changing these parameters by a factor of 100 has almost no effect on the simulation. These parameters are essentially unidentifiable from measurements of species A and R. The right column of Fig. 6 shows how the reduced model changes with the parameter $\bar{\beta}_A$. In the reduced model, changing $\bar{\beta}_A$ by a small amount has a large effect on the simulation, and this single parameter is easily estimated from measurements of A and R.

An additional benefit of the sQSPA model reduction of the biochemical oscillator is that the computation time is re-

Downloaded 17 May 2011 to 131.215.220.185. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

duced to 160 fold. This reduction in computation time occurs because sQSPA reduced model removes the last two reactions of the reaction set (15). Reaction set (15) shows that formation of A from D_A or D'_A occurs through M_A. SSA simulation of the full model spends significant amount of time executing the fast reactions: degradation of M_A or formation of A from M_A. The sQSPA reduced model does not contain these two reactions and the formation of A occurs directly from D_A or D'_A, which results in significant computational savings.

Example 3: Fast fluctuation. In a stochastic simulation of the infection cycle of vesicular stomatitis virus (VSV), fast fluctuation in a protein at low copy number along with a rapid increase in the population of the viral genome occurs. Such a system is expensive to simulate because the frequency of the fluctuation increases as the simulation progresses leading to small time steps in the SSA simulation.[12] To illustrate the phenomenon, consider the following simple three-species, three-reaction system:

$$\mathrm{A} + \mathrm{G} \xrightarrow{k_1} \mathrm{C} + \mathrm{G}, \tag{20}$$

$$\mathrm{C} + \mathrm{G} \xrightarrow{k_2} 2\mathrm{G} + \mathrm{A}, \tag{21}$$

$$2\mathrm{G} \xrightarrow{k_3} \mathrm{G}, \tag{22}$$

$$\text{with } k_2 \sim k_1 \gg k_3.$$

This reaction system describes the interaction of three species in a simplified VSV replication process—two forms of viral polymerase, A and C, and viral genome G. The two forms of polymerase are motivated by the fact that VSV has two different complexes that serve as viral transcriptase and replicase.[28] The viral transcriptase form A is a complex of constituent VSV proteins L and P. The replicase form C is a complex of L, N, and P proteins. A is involved in the transcription reaction (20) to produce messenger RNA. The transcription reaction leads to the conversion of transcriptase A into replicase C. We further assume that produced mRNA from reaction (20) is short lived and hence we do not include it in the model. Species C and G are involved in replication reaction (21) to produce an additional viral genome G. The replication (21) reaction leads to the conversion of replicase C into transcriptase A. Finally, there is a second-order degradation reaction (22) of viral genome. The model (20)–(22) is insufficient to predict the full viral infection cycle, but it is instructive in understanding the simulation challenges of the full infection cycle model used by Hensel *et al.*[12] The reaction rate constants k_1, k_2, k_3 denote macroscopic reaction rate constants with units $\mu\mathrm{m}^3/(\mathrm{mol/s})$. We express microscopic reaction rates in terms of macroscopic rate constants (k_1, k_2, k_3) and the system size Ω

$$r_1 = \frac{1}{\Omega}k_1 ag \qquad r_2 = \frac{1}{\Omega}k_2 cg \qquad r_3 = \frac{1}{\Omega}k_3 g(g-1),$$

in which the system size appears because the reactions are second order. For the purposes of this example we take $\Omega = 10^5 \mu\mathrm{m}^3$. A stochastic simulation with the parameter values given in Table IV is shown in the first row of Fig. 7. Species G increases continuously and this increase forces species C to fluctuate with increasing frequency as shown in the first row of Fig. 7. The left side of Fig. 8 shows that as the simulation progresses, the time step decreases to small values (10^{-6} to 10^{-12}s), which indicates a significant slow down of the SSA simulation. To generate the left side of Fig. 8, we have selected a few time points and plotted the time steps of SSA (Δt) at those time points. The right side of Fig. 8 shows that the frequency of the fluctuation of species C increases with time. To calculate the frequency of fluctuation, we select a few equidistant time points (0.25, 0.75, 1.25, . . .) s and a 0.5 s interval containing each of the time points. We count the number of times C changes during these time intervals. The frequency of the fluctuation at a time point is obtained by dividing the number of times C changes by the length of the interval. Both the left and the right side of Fig. 8 demonstrate the slow down of SSA as simulation progresses due to the rapid increase in species G.

TABLE IV. Initial population and reaction rate constants for the fast fluctuation example.

Species	initial number	Rate constant	value ($\mu\mathrm{m}^3/(\mathrm{mol/s})$)
A	3	k_1	9×10^5
C	0	k_2	5×10^5
G	1	k_3	5×10^{-2}

To reduce this model, we apply the sQSPA-Ω technique in which we first represent G as a continuous variable

$$G = \Omega\phi_G + \Omega^{1/2}\xi,$$

in which ϕ_G is the deterministic mean and ξ is the noise. We substitute this expression into the master equation (20)–(22) to obtain a transformed master equation. Then we collect terms of the same order in Ω. As shown in the supporting information, the $O(\Omega^0)$ terms lead to an approximate evolution equation for species A and C. This evolution equation is a

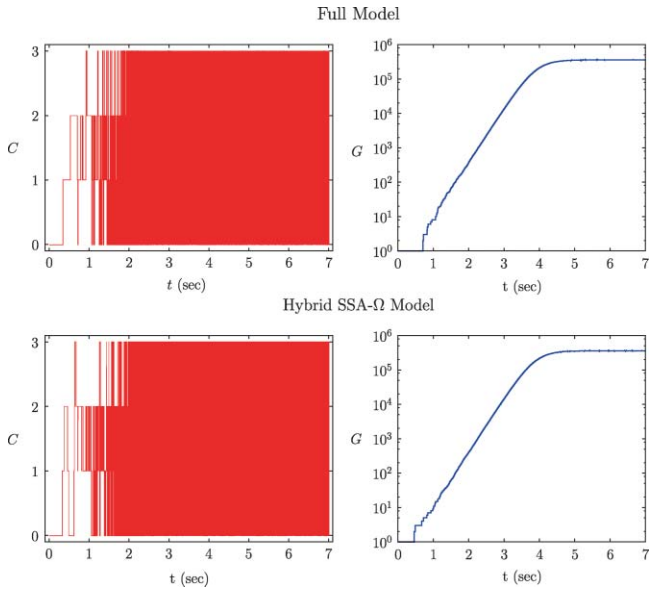

FIG. 7. Top row: species C and G vs time from full model simulation. Bottom row: species C and G versus time from Hybrid SSA-Ω model simulation.

Downloaded 17 May 2011 to 131.215.220.185. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

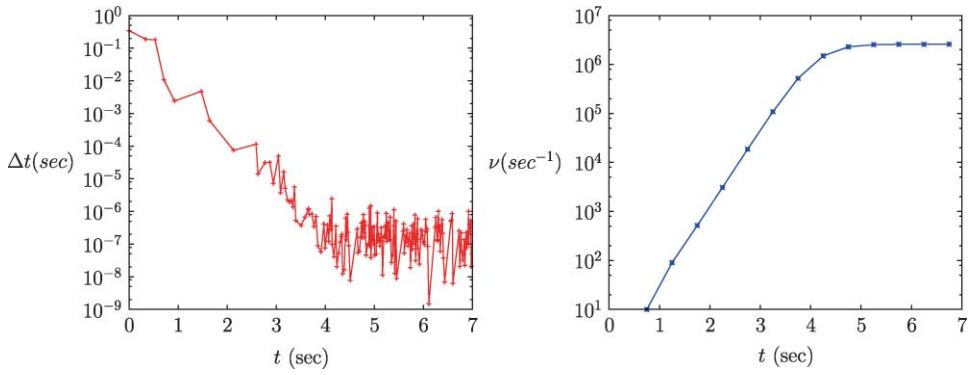

FIG. 8. Left: The step size (Δt) of the SSA vs time; small time steps indicate slow simulation. Right: Frequency (ν) of fluctuation of C vs time.

time-invariant binomial distribution

$$W_{0a}(a) = \binom{N_0}{a}\left(\frac{1}{K_1\gamma+1}\right)^a\left(\frac{K_1\gamma}{K_1\gamma+1}\right)^{N_0-a}, \quad (23)$$

$$W_{0c}(c) = \binom{N_0}{c}\left(\frac{K_1\gamma}{K_1\gamma+1}\right)^c\left(\frac{1}{K_1\gamma+1}\right)^{N_0-c}, \quad (24)$$

in which $\gamma = \Omega/k_2$.

Collecting $O(\Omega^{-1/2})$ terms leads to an evolution equation for ϕ_G

$$\frac{d\phi_G}{dt} = -\gamma^{-1}\langle c\rangle\phi_G + k_3\phi_G^2, \quad (25)$$

where $\langle c\rangle = (N_0K_1\gamma)/(K_1\gamma+1)$ is the mean of the population of species C. Equation (25) shows the dependence of the evolution of the mean concentration of G, ϕ_G on the mean of population of species C, $\langle c\rangle$. Collecting the $O(\Omega^{-1})$ terms give an evolution equation for the probability density of the noise in G, ξ

$$\frac{\partial W_{0\xi}(\xi)}{\partial t} = -\frac{\partial}{\partial\xi}\left[\left(\frac{\langle c\rangle}{\gamma} - 2k_3\phi_G\right)\xi\, W_{0\xi}(\xi)\right] + \frac{1}{2}\frac{\partial^2}{\partial\xi^2}\left[\left(\frac{\langle c\rangle\phi_G}{\gamma} + k_3\phi_G^2\right)W_{0\xi}(\xi)\right]. \quad (26)$$

This evolution equation for $W_{0\xi}(\xi, t)$ is a linear Fokker–Planck equation with time varying coefficients. The equivalent stochastic differential equation describing this noise process is[29]

$$d\xi = \left(\frac{\langle c\rangle}{\gamma} - 2k_3\phi_G\right)\xi\, dt + \sqrt{\frac{\langle c\rangle\phi_G}{\gamma} + k_3\phi_G^2}\; dW, \quad (27)$$

in which W is the continuous time Wiener process, or integrated white noise. Equations (25) and (27) together characterize the population of species G.

Initially when the population of G is small, we perform full SSA for the system. After the onset of the fast fluctuation, we switch to the description of the system given by Eqs. (23)–(27). The onset of the fast fluctuation is determined by setting a threshold value for G. We choose 10^5 for the threshold in this example. We call this combined approach hybrid SSA-Ω, which uses SSA initially and sQSPA-Ω after the onset of the fast fluctuation. Figure 7 shows a comparison of the full SSA and the hybrid SSA-Ω simulations. The left column of Fig. 7 has a sample evolution of C from full SSA (top row) and hybrid SSA-Ω (bottom row), respectively. The right column

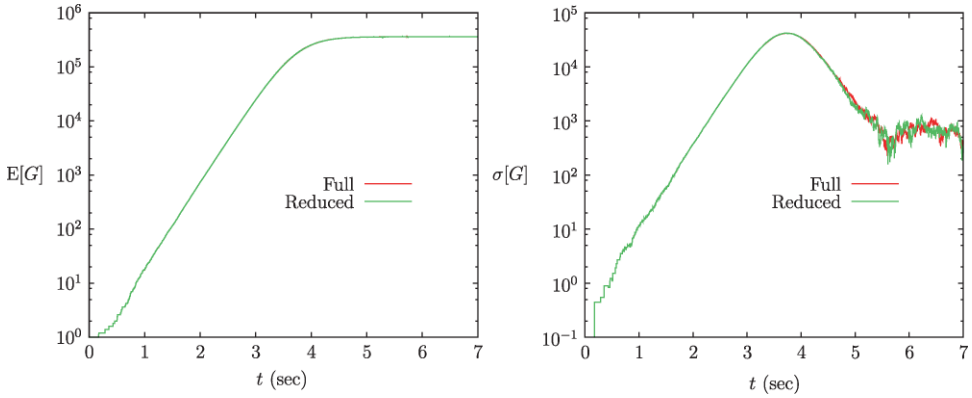

FIG. 9. Left side: Mean of G from full model and Hybrid SSA-Ω reduced models using 500 SSA simulations. Right side: Standard deviation of G from full model and Hybrid SSA-Ω reduced models using 500 simulations.

Downloaded 17 May 2011 to 131.215.220.185. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

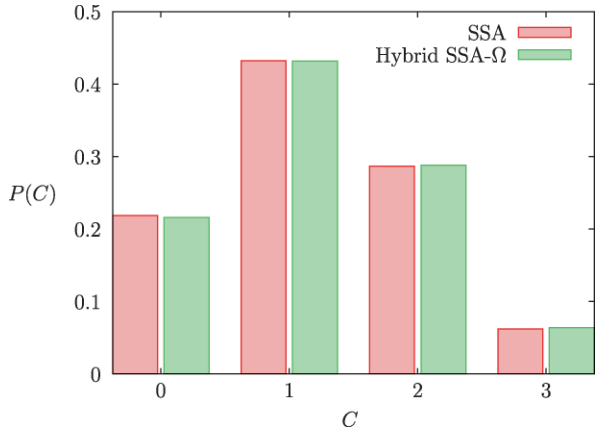

FIG. 10. Comparison of probability density of C obtained from full SSA and Hybrid SSA-Ω.

of Fig. 7 shows a sample evolution of G from the full SSA (top row) and the hybrid SSA-Ω (bottom row). We see that the evolution of C and G from the two simulation approaches match closely. We next compare statistics of G from full and hybrid SSA-Ω simulations. The left side of Fig. 9 shows the mean of G from the full simulation and from the hybrid SSA-Ω simulation, and we see that the hybrid SSA-Ω accurately captures the mean of G. The right side of Fig. 9 shows the same comparison for the standard deviation of G, and we see that the hybrid SSA-Ω also accurately captures the standard deviation of G. Figure 10 compares probability density of C in the full model to the hybrid SSA-Ω model. To obtain an estimate of the probability density of C from the full SSA, we take simulation data from one full SSA simulation after the switch to the hybrid SSA-Ω has occurred. With this simulation data we obtain the frequency distribution of the C species. We take this frequency distribution as an estimate of the probability density of C from the full SSA. The probability density of C from the hybrid SSA-Ω is obtained using Eq. (24). The two densities match closely indicating that the hybrid SSA-Ω model captures the dynamics of the fast fluctuating species accurately.

Comparison of the simulation times for the two models shows an 80-fold reduction with hybrid SSA-Ω. A single simulation of SSA requires about 111 s whereas a single simulation of hybrid SSA-Ω requires 1.45 s.[30]

III. CONCLUSIONS

This paper demonstrates the usefulness of the stochastic quasi-steady-state singular perturbation analysis and the stochastic quasi-steady-state singular perturbation analysis with Ω expansion (sQSPA-Ω).[23] These methods are well suited for situations in which some of the species in the system are highly reactive and sample mainly small (1−10) values. The reactant and product species may be present in either moderate values ($< 10^3$), in that case sQSPA is suitable, or in large values ($\gg 10^3$), in that case sQSPA-Ω is suitable.

We considered three examples here. The first two examples, *pap operon regulation* and a *biochemical oscillator*, use sQSPA and the third one, *fast fluctuation*, uses sQSPA-Ω. The *pap operon regulation* full model consisted of four states, and the sQSPA reduction removed one low probability state (g_1). This example highlights difficulties in accurate rare state probability estimation using either SSA simulations or the recent technique proposed by Kuwahara and Mura based on importance sampling.[24] We showed that accurate estimation of the rare state probability (g_1) can be obtained using the probabilities of non-rare (g_2, g_3) states and the analysis provided by sQSPA. The *biochemical oscillator* example contained one species (M_A) that sampled mainly zero because of the rapid M_A degradation reactions. The sQSPA reduction converted the estimation problem of two large rate constants into the estimation of one well scaled parameter, their ratio. Also the reduced model showed a 160-fold decrease in simulation time. The *fast fluctuation* problem exhibited one species fluctuating rapidly and another increasing rapidly. We applied an extension of sQSPA-Ω that we call hybrid SSA-Ω. Hybrid SSA-Ω leads to an approximate probability distribution of the fast fluctuating species. It leads to a combination of an ordinary differential equation and a Langevin equation for the rapidly increasing species. The reduced model of the *fast fluctuation* problem showed an 80-fold speed up in the computation time.

In stochastic modeling of biological systems, certain characteristics are desirable: fast simulation times, ability to compute rare state probabilities directly from simulation, and ability to readily estimate model parameters from experimental data. Many detailed biological models do not have some or any of these characteristics. Model reduction tools such as sQSPA and sQSPA-Ω can play useful roles in developing reduced models that do have these characteristics.

ACKNOWLEDGMENTS

This work was supported by a National Institutes of Health (AI071197) award. E.L.H. gratefully acknowledges support from the National Institutes of Health under Ruth L. Kirschstein National Research Service Award 5F32CA120055. All simulations were performed using Octave (http://www.octave.org). Octave is freely distributed under the terms of the GNU General Public License.

[1] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Science **297**, 1183 (2002).
[2] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, Nat. Genet. **31**, 69 (2002).
[3] W. J. Blake, M. A. Kaern, C. R. Cantor, and J. J. Collins, Nature (London) **422**, 633 (2003).
[4] J. M. Raser and E. K. O'Shea, Science **304**, 1811 (2004).
[5] S. Hooshangi, S. Thiberge, and R. Weiss, Proc. Natl. Acad. Sci. U.S.A. **102**, 3581 (2005).
[6] B. Munsky and M. Khammash, J. Chem. Phys. **124**, 044104 (2006).
[7] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[8] D. T. Gillespie, Physica A **188**, 404 (1992).
[9] A. Arkin, J. Ross, and H. McAdams, Genetics **149**, 1633 (1998).
[10] J. M.G. Vilar, H. Y. Kueh, N. Barkai, and S. Leibler, Proc. Natl. Acad. Sci. U.S.A. **99**, 5988 (2002).
[11] L. S. Weinberger, J. C. Burnett, J. E. Toettcher, A. P. Arkin, and D. V. Schaffer, Cell **122**, 169 (2005).
[12] S. Hensel, J. B. Rawlings, and J. Yin, Bull. Math. Biol. **71**, 1671 (2009).
[13] D. T. Gillespie, Annu. Rev. Phys. Chem. **58**, 35 (2007).
[14] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna, PLOS Comput. Biol. **3**, 1871 (2007).

Downloaded 17 May 2011 to 131.215.220.185. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

[15]Y. Cao, D. T. Gillespie, and L. R. Petzold, J. Chem. Phys. **122**, 014116 (2005).
[16]J. Goutsias, J. Chem. Phys. **122**, 184102 (2005).
[17]E. L. Haseltine and J. B. Rawlings, J. Chem. Phys. **123**, 164115 (2005).
[18]A. Samant and D. G. Vlachos, J. Chem. Phys. **123**, 144114 (2005).
[19]H. Salis and Y. Kaznessis, J. Chem. Phys. **122**, 054103 (2005).
[20]W. E. D. Liu and E. Vanden-Eijnden, J. Chem. Phys. **123**, 194107 (2005).
[21]W. E. D. Liu and E. Vanden-Eijnden, J. Comput. Phys. **221**, 158 (2007).
[22]C. V. Rao and A. P. Arkin, J. Chem. Phys. **118**, 4999 (2003).
[23]E. A. Mastny, E. L. Haseltine, and J. B. Rawlings, J. Chem. Phys. **127**, 094106 (2007).
[24]H. Kuwahara and I. Mura, J. Comput. Phys. **129**, 165101 (2008).
[25]See supplementary material at http://dx.doi.org/10.1063/1.3580292 for the proof.
[26]M. Roh, D. Gillespie, and L. Petzold, J. Comp. Phys. **133**, 174106 (2010).
[27]Note 1, note that $\bar{\delta}_{MA} = 1 - \bar{\beta}_A$, which means that $\bar{\delta}_{MA}$ is not an independent parameter in the reduced model.
[28]J. K. Rose and M. A. Whitt, in *Fields Virology*, edited by D. M. Knipe and P. M. Howley (Lippincot Williams & Wilkins, Philadelphia, 2001), vol. 1, pp. 1221–1244, 4th ed.
[29]P. E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations* (Springer Verlag, Berlin, 1992).
[30]Note 2, all computations were done with Octave on AMD Athlon Dual core processor with clock speed 2.2 GHz, L1 cache memory 256KB and L2 cache memory 2MB.

Downloaded 17 May 2011 to 131.215.220.185. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions